4.3 Numeric covariate

13

$z_i \in \mathcal{Z} \subseteq \mathbb{R}$ numeric

- sensible (and often used) choices of reparameterization will follow

4.3.1 Simple transformation of the covariate

14

$$\boxed{m(z) = \beta_0 + \beta_1 s(z)}, \quad z \in \mathcal{Z}$$

$s: \mathcal{Z} \to \mathbb{R}$ non-constant function

e.g. $s(z) = \log(z)$

$$\Rightarrow m(z) = \beta_0 + \beta_1 \log(z)$$

$$X = \begin{pmatrix} 1 & \log(z_1) \\ \vdots & \vdots \\ 1 & \log(z_n) \end{pmatrix}$$

HERE: if $Z \sim$ continuous distribution

$P(\text{rank}(X_{n \times 2}) = 2) = 1$ (if $n \geq 2$)

Example: $Y = \log(\text{price})$, Z = ground size 15

$X := \log(z)$

regression function: line in $x = s(z)$

$\mathbb{E}(Y \mid Z = z)$

$\mathbb{E}(Y | Z = z) = \beta_0 + \beta_1 \cdot \log(z)$ 16

$\beta_1 = \mathbb{E}(Y | \underbrace{\log(Z)}_{s(Z)} = x+1) - \mathbb{E}(Y | \underbrace{\log(Z)}_{s(Z)} = x)$

here: $\hat{\beta}_1 = 0{,}54$

change of $\mathbb{E}(Y|\ldots)$ corresponding to unity increase of $s(Z)$ and not Z itself

---

assumptions? 17

Does Z influence $\mathbb{E}(Y|\cdots)$? 18

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

R output (t-test) 19

4.3.2 Raw polynomials

20

$$\boxed{m(z) = \beta_0 + \beta_1 z + \dots + \beta_{k-1} z^{k-1}}, \quad z \in \mathcal{Z}$$

$$s(z) = (z, z^2, \dots, z^{k-1})^T$$

$$X = \begin{pmatrix} 1 & \underbrace{\begin{matrix} z_1 & \dots & z_1^{k-1} \\ \vdots & & \vdots \\ z_n & \dots & z_n^{k-1} \end{matrix}}_{S} \\ \vdots & \\ 1 & \end{pmatrix}$$

- $\underline{k}$ different values among z_i's
  $\Rightarrow \text{rank}(X_{n \times k}) = k$
- if $Z \sim$ continuous distribution
  $P(\text{rank}(X_{n \times k}) = k) = 1$ (if $n \geq k$)

Example: $Y = \log(\text{price})$, $Z =$ ground size 21

---

assumptions? 22

---

cubic polynomial 23-24

Does Z influence $E(Y/\dots)$? 25

$H_0: \beta^Z = 0$ $\quad (\beta^Z = (\beta_1, \dots, \beta_{k-1})^T)$

$H_1: \beta^Z \neq 0$

- Wald type test (F-test)
- submodel F-test

Interpretation of β's ? 26

R output 27

Lower degree $(d-1)$ sufficient $(d<k)$ 28
to express the regression function ?

$H_0: \beta_d = 0, \dots, \beta_{k-1} = 0$

- Wald type test (F-test)
- submodel F-test

R output 29

Graphical comparison 30
log vs. polynomial.

Assumptions? 31

Practical importance of higher 32
order polynomial terms?

- be aware of a scale

4.3.3 Orthonormal polynomials

→ equivalent linear model to <u>raw</u> polynomials

$$\boxed{m(z) = \beta_0 + \beta_1 P^1(z) + \dots + \beta_{k-1} P^{k-1}(z)}$$

$$P^j(z) = a_{j,0} + a_{j,1} z + \dots + a_{j,j} z^j \quad , j = 1, \dots, k-1$$

- <u>orthonormal polynomial</u> of degree j built above a set of the covariate datapoints $z_1, \dots, z_n$

$$X = \begin{pmatrix} 1 & P^1(z_1) & \dots & P^{k-1}(z_1) \\ \vdots & \vdots & & \vdots \\ 1 & P^1(z_n) & & P^{k-1}(z_n) \end{pmatrix}$$

$$= \left(\mathbb{1} \quad \underbrace{\mathbb{P}^1 \quad \dots \quad \mathbb{P}^{k-1}}_{\mathbb{S}} \right)$$

- coefficients ~~are~~ $a_{j,l}$ are (and can be (if at least k different values among $z_1, \dots, z_n$)) chosen such that

- $\mathbb{P}^1, \dots, \mathbb{P}^{k-1}$ are mutually orthogonal
- $\mathbb{P}^j \perp \mathbb{1}$
- $\|\mathbb{P}^j\| = 1$ → same <u>scale</u> for all cols. of X

 one more useful property: $X^T X \overset{\text{diag}}{=} (n, 1, \dots, 1)$

R output 34

Fitted line (cubic polynomials) 35

Assumptions? 36

Basis polynomials (plot) 37

Compare $\hat{\beta}$'s 38

Does Z influence $E(Y|\cdots)$? 39

- the same as with raw polyn.

R output 40

Interpretation? 41

Lower degree sufficient? 42

- the same as with raw polyn.

R output 43-44

Global influence of each datapoint 45

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_i (Y_i - \beta_0 - \beta_1 z_i - \beta_2 z_i^2 - \beta_3 z_i^3)^2$$

4.3.4 Regression splines

Def 4.2 Basis spline with distinct knots

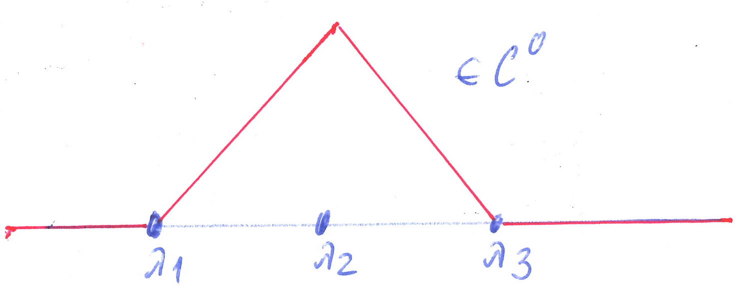
Let $d \in \mathbb{N}_0$ and $\lambda = (\lambda_1, \dots, \lambda_{d+2})^T \in \mathbb{R}^{d+2}$, $-\infty < \lambda_1 < \dots < \lambda_{d+2} < \infty$. The basis spline of degree d with distinct knots λ is such a function $B^d(z;\lambda)$, $z \in \mathbb{R}$ that

(i) $B^d(z;\lambda) = 0$, $z \le \lambda_1$, $z \ge \lambda_{d+2}$;

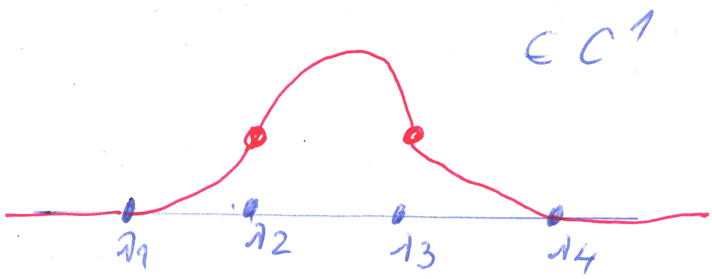
(ii) On each of the intervals $(\lambda_j, \lambda_{j+1})$, $j = 1, \dots, d+1$, $B^d(\cdot;\lambda)$ is a polynomial of degree d;

(iii) $B^d(\cdot;\lambda)$ has continuous derivatives up to an order $d-1$ on $\mathbb{R}$

$d=1$ (linear)



$d=2$ (quadratic)



– plot

→ TO REMEMBER:

Each $B^d(\cdot;\lambda)$ is

- piecewise polynomial
- smooth ($\in C^{d-1}$ at inner knots, with 0 on boundary)

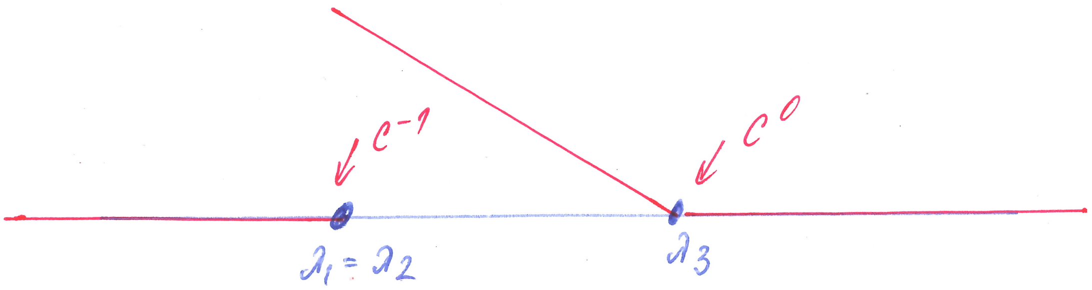
Def 4.3 Basis spline with coincident left boundary knots [48]

Let $d \in \mathbb{N}_0$, $1 < r < d+2$ and $\lambda = (\lambda_1, \dots, \lambda_{d+2}) \in \mathbb{R}^{d+2}$,

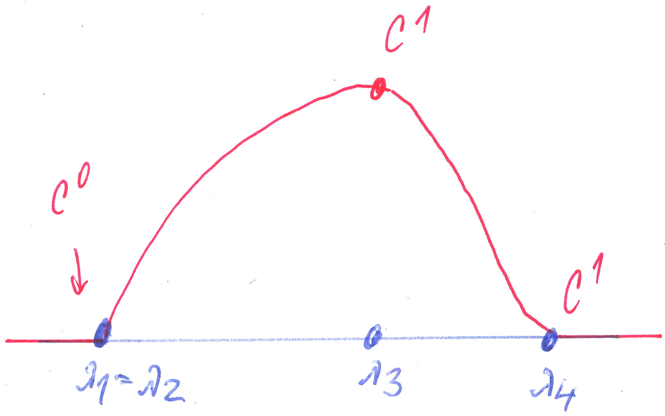
$$-\infty < \lambda_1 = \dots = \lambda_r < \dots < \lambda_{d+2} < \infty .$$

etc. . .

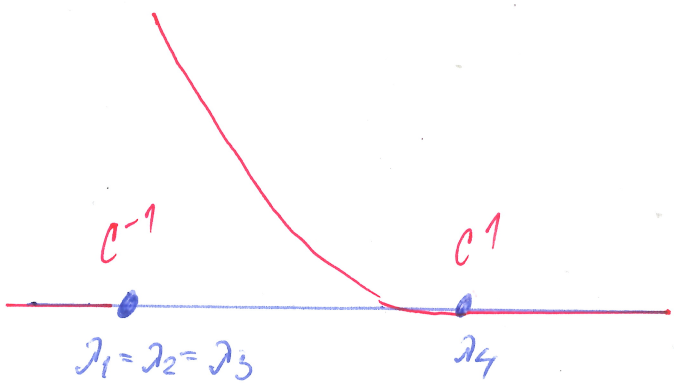
$d=1$, $r=2$



$d=2$, $r=2$



$r=3$



mirror definition : Basis spline with coincident right boundary knots

plots [49-51]

There exist many ways on how to construct splines that satisfy stated definitions. 52

Previous plots showed basis B-splines.

Properties useful for regression modelling:

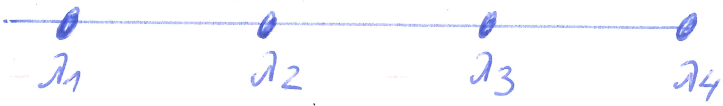
- $B^d(z;\lambda) > 0$, $\lambda_1 < z < \lambda_{d+2}$
- $B^d(z;\lambda) = 0$, $z \le \lambda_1$, $z \ge \lambda_{d+2}$
- smoothness

Def 4.4 Spline basis 53

. . . . .

$d=2,\ k=5$ $d=2$: basis spline needs **4 knots**

$\downarrow k-d+1=4$



B_1: $\lambda_1=\lambda_1=\lambda_1,\ \lambda_2$

B_2: $\lambda_1=\lambda_1,\ \lambda_2,\ \lambda_3$

B_3: $\lambda_1,\ \lambda_2,\ \lambda_3,\ \lambda_4$

B_4: $\lambda_2,\ \lambda_3,\ \lambda_4=\lambda_4$

B_5: $\lambda_3,\ \lambda_4=\lambda_4=\lambda_4$

---

spline basis $(d=1,2,3)$ 54-56

Properties of the B-spline basis (again useful for statistical modelling) 57

(a) $\sum_{j=1}^{k} B_j(z) = 1 \quad \forall z \in (\lambda_1, \lambda_{k-d+1})$.

(b) for each $m \le d$ there exist a set of coefficients $\gamma_1^m, \ldots, \gamma_k^m$ such that $\sum_{j=1}^{k} \gamma_j^m B_j(z)$ is on $(\lambda_1, \lambda_{k-d+1})$ a polynomial in z of degree m.

Regression spline 58

Covariate space $\mathcal{Z} = (z_{min}, z_{max})$, $-\infty < z_{min} < z_{max} < \infty$

→ choose $d \in \mathbb{N}_0$, set of knots $\lambda = (\lambda_1, \ldots, \lambda_{k-d+1})^T$,
$z_{min} = \lambda_1$, $z_{max} = \lambda_{k-d+1}$

→ spline basis $B_1(z), \ldots, B_k(z)$

→ regression function ≡ MODEL FOR $\mathbb{E}(Y | Z = z)$:

$$\boxed{m(z) = \beta_1 B_1(z) + \ldots + \beta_k B_k(z), \quad z \in \mathcal{Z}}$$

$\beta = (\beta_1, \ldots, \beta_k)^T$: (unknown) regression coefficients

DATA: $(Y_1, Z_1)^T, \ldots, (Y_n, Z_n)^T$

→ reparameterizing matrix ≡ model matrix

$$X = \mathbb{S} = \begin{pmatrix} B_1(Z_1) & \ldots & B_k(Z_1) \\ \vdots & & \vdots \\ B_1(Z_n) & \ldots & B_k(Z_n) \end{pmatrix}$$

$$X = \mathbb{S} = \begin{pmatrix} B_1(z_1) & \cdots & B_k(z_1) \\ \vdots & & \vdots \\ B_1(z_n) & \cdots & B_k(z_n) \end{pmatrix}$$

$$= \left(\mathbb{B}^1, \ldots, \mathbb{B}^k\right) =: \mathbb{B}$$

TO REMEMBER:

- $\forall z \in (\lambda_1, \lambda_{k-d+1}) \quad \sum_{j=1}^{k} B_j(z) = 1$

  $\Rightarrow \mathbb{1} \in \mathcal{M}(\mathbb{B})$

- $\forall m \leq d \quad \exists \gamma_1^m, \ldots, \gamma_k^m : \sum_{j=1}^{k} \gamma_j^m B_j(z)$

  is a (global) polynomial of degree m on $(\lambda_1, \lambda_{k-d+1})$

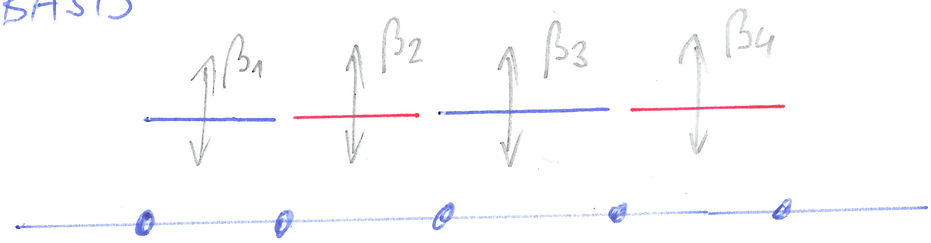
  $\Rightarrow \mathcal{M}(\mathbb{P}^m) \subset \mathcal{M}(\mathbb{B})$

  $\mathbb{P}^m \equiv$ model matrix based on raw/orthonormal polynomials of degree m

$\Rightarrow$ submodel F-test can be used to test H_0: regression function is a (global) polynomial of degree m

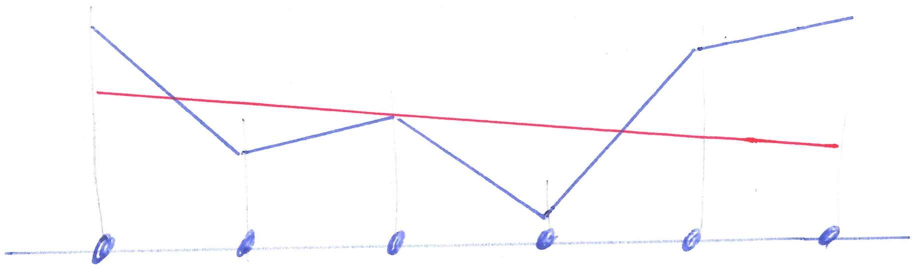
REMARK:

- $\underline{\underline{d=0}}$ BASIS



regression function $\equiv$ piecewise constant

- $\underline{\underline{d=1}}$



- regression function $\equiv$ piecewise linear (& continuous)
- we can test H_0: $m(z)$ is LINEAR on (ν_1, ν_{k-d+1})

- $\underline{d=2,3,\dots}$
  - regression function is piecewise quadratic, cubic, ... (& smooth)

IMPORTANT ISSUE IN PRACTICE:

- number of knots
- placement of knots

→ out of the scope of this lecture

→ trial/error methods applied here, but more sophisticated procedures exist

Example: Houses 1987 [59]

- spline basis ($d=3$)

5 knots $\to k=7$

B matrix [60]

$$\begin{pmatrix} B_1(z_1) & \cdots & B_k(z_1) \\ \vdots & & \vdots \\ B_1(z_n) & \cdots & B_k(z_n) \end{pmatrix}$$

R output [61]

R^2, F to be highligted

F-test: $H_0: \beta_1 = \cdots = \beta_7 = 0$

Fitted function (+ prediction band) [62]

Diagnostics [63-64]

$H_0: m(z) = \text{const}$ [65]

$\equiv \beta_1 = \cdots = \beta_7$

$\equiv E(Y|Z) \in \mathcal{M}(\mathbb{1})$ [66] $\to$ F-test on submodel

---

$H_0: m(z) = \text{cubic polynomial}$ [67]

$\equiv E(Y|Z) \in \mathcal{M}(\mathcal{P}^3)$

$\to$ F-test on submodel

Comparisons. [68]

Interpretation of β's? [69]

"Real" modelling.

Motorcycle (helmet) dataset.

[70-75]

Reminder from matrix calculus, partly needed next time!

<u>Kronecker product</u> of two matrices:

$A_{m\times n}$, $C_{p\times q}$:

$$A \otimes C = \begin{pmatrix} a_{11} C & \cdots & a_{1,n} C \\ \vdots & & \vdots \\ a_{m,1} C & \cdots & a_{m,n} C \end{pmatrix}$$

$a \in \mathbb{R}^m$, $c \in \mathbb{R}^p$

$$a \cdot c^T = a \otimes c^T$$